

The Machine Learning Model

- We have chosen to use oversampling and undersampling as our machine learning model
- Oversampling is the potential duplication of data within a dataset
- Undersampling is the potential under-representation of data within a dataset

Understanding the Machine Learning Model

- For our topic we have chosen to utilize an oversampling algorithm and an undersampling algorithm to predict the correlation between oil prices and the impact on the US dollar. We will then use the same algorithms to analyze how world events will impact oil prices and the strength of the dollar
- When we combine the algorithms, the final result will predict the oil price and present an image to show the correlation and the accuracy of the models.
- The oversampling and undersampling algorithms used in this project can be classified as unsupervised machine learning models because no labels will be assigned to the data. The data will be examined for a correlation using a best fit model.

Benefits of this Model for Oil Data Analysis

- Oversampling and undersampling involve introducing a bias to into the oil and dollar value data to help compensate for an imbalance hidden within the data set.
- Oversampling and undersampling are commonly referred to as “*naive resampling*” methods because they assume nothing about the data and no heuristics are used. This makes them simple to implement and fast to execute, which is desirable for very large and complex datasets such as the ones we are using for this project.

Description of Model's Accuracy

- Oversampling and undersampling do have some drawbacks such as duplicating data which may result in overfitting a best fit line used to show a correlation between oil prices and the dollar value.
- However, we plan on removing duplicated sets of data within our code which will enable us to have an accurate representation of how oil prices and the dollar valued are related

Importance of Statistical Data for Oil and Dollar Value Correlation

- Oversampling and undersampling are important for the correlation because